Effect of GDP Per Capita on National Life Expectancy

Gokce Dayanikli, Varsha Gokare, Brian Kincaid

Abstract

Using cross sectional data from several sources, this paper seeks to find a correlation between per capita GDP, public health expenditure, and average years of education with the cross country life expectancy as measured at birth. By adding the additional factors, the paper shows that there are underlying issues that affect the overall life expectancy throughout the world. A better understanding, while not all inclusive, will help public officials, aide organizations, and world health policy makers more effectively appropriate resources in an effort to increase global life expectancy.

1. Introduction

Humanity, throughout history, has made a connection between the possession of material goods and happiness. This is a connection, that if true, implies that simply possessing wealth can make one better off. Is this connection present, and is there a direct connection between wealth and happiness? While happiness cannot be easily monitored, measured, or categorized thus making a direct comparison difficult at best; other aspects of human life can more readily be quantified.

As an analog for happiness, many attributes of have been used including lifespan, birthrates, education, spending habits as well as countless others. For our research purposes, we have focused on mortality as an easily measurable quantity to draw conclusions around. Does a higher GDP lead to lower mortality rates or is there some subtler set of details working to connect income levels and death rates?

We sought to draw connections between a country's GDP and mortality rates in that country, as well as to draw new intermediate and tangential connections that may link GDP and mortality. We also sought to find out if there is a linear correlation between mortality rates for a certain country and that country's GDP, or are the other factors creating a non-linear relationship.

Many research projects have sought to draw correlation, even causality, between GDP and mortality. Any correlation is quantifiable evidence that can be used to improve arguments that money can improve, even at a most basic level, the quality of life in a country. However, it cannot simply direct how that money should be spent. If a more targeted approach for spending could be found, perhaps on infrastructure, transportation, or education then perhaps countries with lower GDPs would be able to improve mortality rates without often large and impossible increases to their population's income level.

2. Literature Review

A great deal of research has been performed on whether wealth can, in fact, increase happiness. Among the many papers, projects, surveys, and proposals are hypotheses showing varying levels of correlation between per capita GDP and quality of life. These also include several works by recent Nobel laureate, Sir Angus Deaton, a professor Princeton University. One of these works, titled "Health in an Age of Globalization" (2004) seeks to debunk previous theories that globalization has increased death rates in countries that have increased GDP through trade. In this paper, Deaton is able to draw several conclusions regarding the correlation between trade, per capita GDP, and mortality rates around the world.

The first and perhaps most important correlation that is shown, is that while increased levels of trade between countries of the world has led to an increase of transmission of communicable disease, the corresponding transfer of medical knowledge has led to an increase in life expectancy among countries with similar GDP. The same diseases that are being moved from continent to continent are being cured or rendered non-life-threatening.

While Deaton identifies that per capita GDP does have a correlation with life expectancy, the research focuses more on income inequality as being the primary barrier to reducing mortality rates in poor countries. Throughout the paper, however, Deaton does make exception to the spread of HIV/AIDS in sub-Saharan Africa as an outlier to the data. The correlation between GDP and life expectancy also begins to breakdown after a certain threshold is reached in per capita GDP. After that point, the original linear regression no longer fits the data.

Deaton suggests, as the data in our research also shows, that while per capita GDP can affect the lifespan of a population within a country, it may not be entirely due to the actual amount spent but rather how it is spent. Improvements in infrastructure, technology, and education can not only explain the improvement among low GDP countries, but can also help explain why higher GDP countries don’t follow the linear increase in lifespan that accompanies low GDP countries.

Dr. Gary Becker, an economist at the University of Chicago, also released a paper showing the gap inaccuracy of current models that show the correlation between per capita GDP and life expectancy (2005). While Becker’s paper focuses on the lack of accuracy in the model, the reason stated for such variance is not entirely in concert with Deaton’s research. The major gap, as Becker found, comes from income inequality across countries rather than differences in how GDP is spent.

While Becker and Deaton align with one another that HIV/AIDS in Africa is an outlier that has moved the data and the model away from the previously accepted analysis, Becker also introduces several other reasons that may explain why the model grows less accurate as inequality increases. In recent years, Becker suggests, the rapid increase in per capita GDP in China and India have may have skewed the available data to screen other correlations and causations that are present.

A major similarity in the research between the two papers, however, follows standard economic theories that the largest increases come from the countries that start at the lowest GDP. As stated in Becker’s research, from 1960 to 2000 the fastest increase in welfare was seen in the poorest countries around the world while the countries at the top of the GDP curve saw the slowest increase in welfare for the same time period.

The suggestion that factors other than raw per capita GDP, even those that can be affected by GDP itself, is further reinforced by Becker's conclusion showing greater increases in welfare among the poorest of countries. As countries develop at an early stage, rapid increases in transportation infrastructure and information technology help to accelerate the spread of medical technologies, information, and medicine thus increasing the life expectancy of citizen of that country.

A second paper written by Deaton (2008) begins to draw further conclusions from the already established data. While many previous works aim to correlate GDP per capita with life expectancy, this paper attempts to take the correlation to include life satisfaction.

Using data from the Gallup World Poll, Deaton's findings fit a very similar curve to that of the Preston curve. Low income countries, measured in GDP per capita, have lower overall life satisfaction while those with a higher relative GDP per capita rated their lives as more satisfactory. The degree to which changes in income affected satisfaction was also very similar in shape to the original Preston curve, with a more positive correlation at low GDP and a smaller positive correlation at high GDP.

Also included in Deaton's analysis, is the correlation between per capita GDP and health satisfaction. The conclusions drawn in this respect are similar to those drawn from the overall life satisfaction. Overall health satisfaction increases rapidly at lower GDP levels, and continues to increase at a slower rate as GDP increases. While the findings in Deaton's paper support the overall model that per capita GDP has an effect on life expectancy, they also attempt to take the research and analysis one step further to better understand the underlying mechanics involved.

While Becker, Deaton, and many others associate GDP per capita as a main factor in the determination of life expectancy, others have made different hypotheses to better explain the discrepancies associated with the Preston curve. One such model, published by Moshe Hazan (2012), shows that the main factory in the determination of life expectancy across different countries is education. While Hazan states that increases in education beyond the age of five does not necessarily result in an increase in life expectancy, the model does show a weak positive correlation between increases in education and overall increases in life expectancy. While education may very well lead to increases in life expectancy, there is a strong collinearity between increases in education and increases in per capita GDP, and thus the explanation cannot be labeled as causal.

3. Data

3.1 Independent and Dependent Variables

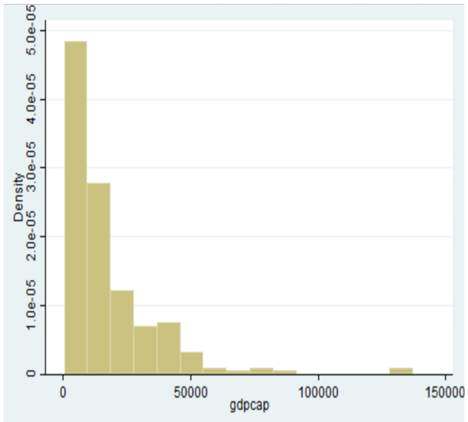
The research was centered around finding a correlation between a nation's GDP per capita (constant 2011 international $), measured with Purchasing Power Parity (PPP) to standardize the data and the nation's life expectancy, in years. For GDP per capita, the one with measured with PPP is used in order to have a relative measure by comparing the standard of living within the countries. The hypothesis was that they have a positive correlation, when the national income increases, the national life expectancy also rises. By this cause-and-effect relationship, GDP per capita is used as the independent variable. GDP per capita is the independent variable because it tells about many macroeconomic indicators about a country: it's wealth, government expenditures, health insurance availability, etc. All of these factors directly invoke a response is life expectancy (our dependent/response variable).

After researching the relationship between GDP and life expectancy, it is further hypothesized that other variables could be explanatory variables. Some other explanatory variables that affect life expectancy are health expenditures as a percent of total government expenditures, information technology, infrastructure, and average schooling years. In addition, everything attributed to GDP per capita (wealth, government expenditures, health insurance availability) can be its own explanatory variable, as long as it doesn't have a perfect correlation with GDP per capita.

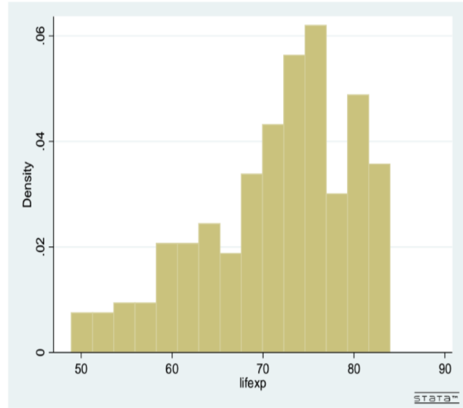
3.2 Source of Data and Descriptive Statistics

All of the research comes from The World Bank, based on the year 2013. The data is found on explanatory and response variables for numerous countries. For GDP per capita (gdpcap), there are 234 observations, with a minimum of 561.46 and a maximum of 137164.4. For life expectancy (lifexp), there are 228 observations, with a minimum of 49 years and a maximum of 84 years. Swaziland has the lowest life expectancy and Hong Kong has the highest life expectancy. The average life expectancy is about 71 years with a standard deviation of 8 years. Note that some of the countries do not have data for both variables, so the total number of observations is 207. All this data can be seen summarized in the Table 1.

In addition, these are the histograms of GDP per capita and life expectancy:

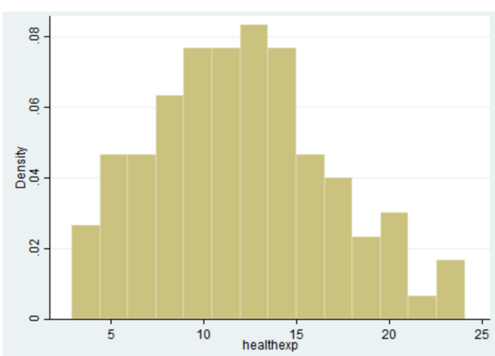


(Figure 1: Histogram of gdpcap)

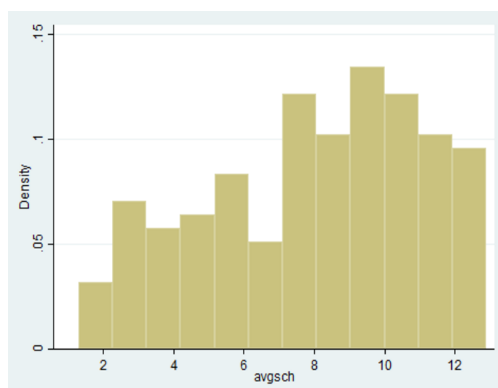


(Figure 2: Histogram of lifexp)

For the multiple regression analysis, another explanatory variable is added: public health expenditure as a percentage of total government expenditure (healthexp). This data also comes from the World Bank, based on year 2013. There are 198 observations, with the mean being about 12% and a standard deviation of about 4.75%. Furthermore, a third explanatory variable is also added: average years of schooling (avgsch). This data was found from the Human Development Report and it is based on year 2013. There are 162 observations, with a minimum of 1.3 years and a maximum of 12.9 years. The mean is about 7.985 years with a standard deviation of 3.098 years. The summary statistics are listed in the Table 1 below, and the histograms of the additional explanatory variables are also seen below.



(Figure 3: Histogram of healthexp)



(Figure 4: Histogram of avgsch)

	# of Observations	Minimum	Maximum	Average	St. Deviation
GDP Per Capita (PPP)	234	561.4612	137164.4	17203.72	19498.16
Public Health Expenditure	198	2.871307	24.05637	12.02605	4.750809
Average Years of Schooling	162	1.3	12.9	7.984568	3.098358
Life Expectancy	244	48.93793	83.83171	70.96474	8.238278

(Table 1: Summary statistics for the variables)

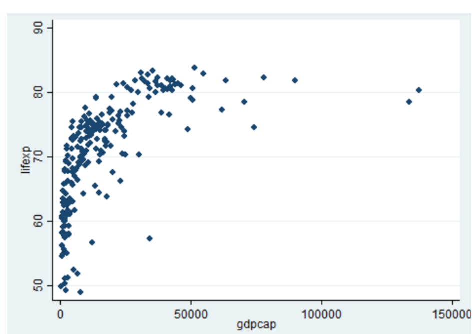
4. Results

4.1 STATA Results for Simple Linear Regression Model and Interpretation of Results

Firstly, a Simple Regression is performed between GDP per capita (national income), PPP and life expectancy.

Regression Model 1: Predicted Life Expectancy at birth = 66.28958 + 0.0002577(GDP per capita)

□1 is found as 0.0002577 . This interpretation is that as the GDP per capita in terms of PPP increases by 1, the life expectancy increases by 0.0002577 years. The prediction of a positive correlation between the two variables is supported by this positive slope.

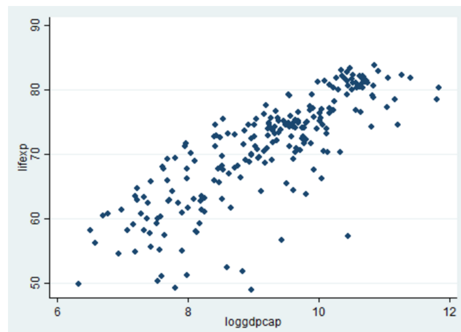


The R-squared value is 37.78%. This interpretation is that 37.78% of the data can be explained by this regression model. Note that 37.78% is not a high R-squared value. Thus, a higher R-squared value can be obtained by increasing the number of explanatory variables. The regression summary can be seen in the STATA output in the Appendix.

However, a curved shape in the data is clearly seen, which explains the low R-squared value. To solve this, a log transformation of GDP per capita was performed. This log transformation drastically straightened out the data and improved the regression model, as seen below. The new R-squared value obtained is 65.43%, which is much higher. This model is more meaningful. Note that □1 changes as well as its interpretation. □1 is found as 5.737 so as GDP per capita increases by *1%*, predicted life expectancy increases by 5.737 years.

Regression Model 2: Predicted Life Expectancy = 18.09359 + 5.737335 x log(GDP per capita)

Although this log transformation drastically improved the regression model, even more could be done. As stated before, the best way to increase R-squared, and the fit of the regression equation, is to add more explanatory variables. A multiple regression analysis is then performed with the second explanatory variable, public health expenditures. This regression is valid because the explanatory variables are not perfectly correlated; they have a correlation of 0.2309. Assumption 3 is met. The R-squared value for this model is 67.35%, which is the highest R-squared thus far.



By adding another explanatory variable, the bias decreases; the expected value of the predicted slope coefficients are closer to the actual, population slope coefficient. The tradeoff is that the variance increases. As R-squared increases, the variance of the betas increase, due to the variance inflation factor.

Regression Model 3: Predicted Life Expectancy = 17.62939 + 5.384642 x log(GDP per capita) + 0.3269995 x public health expenditure

Both of the beta coefficients in this model are positive, which support the hypothesis. Also, this intuitively makes sense. As public health expenditure and GDP per capita increase, the life expectancy so logically increase. To be specific, as GDP per capita increases by 1%, life expectancy increases by 5.38 years; as public health expenditure as a percentage of total government expenditure increases by 1%, life expectancy increases by 0.327 years. Without performing any robustness tests (which will be performed later), this model is meaningful.

In addition to public health expenditure, average schooling is added as a third explanatory variable, to try and increase R-squared even more. So now, life expectancy is being regressed with log(GDP per capita), public health expenditure, and average years of schooling. Again, Assumption 3 is met, which can be seen in the correlation table below, Table 2. The R-squared value is 68.13%, which is the highest R-squared value now, although marginally. Furthermore, the adjusted R-squared is 67.49%, which is still higher than the all other model's R-squared values. The adjusted R-squared value takes into account that adding more variables will increase R-squared, and it "adjusts" R-squared accordingly.

All coefficients match the logical interpretation of what the relationships should be (positive). As GDP per capita increases by 1%, life expectancy increases by 3.844 years. Similarly, as public health expenditure increases by 1% and average schooling increases by 1 year, life expectancy increases by 0.245 years and 0.793 years, respectively. As a result, model 4 is the most meaningful model, and will be used from now on.

Regression Model 4: Predicted Life Expectancy = 26.34697 + 3.843959 x log(GDP per capita) + 0.2454119 x public health expenditure + 0.7925244 x average years of schooling

	loggdpcap	**healthexp**	**avgsch**
loggdpcap	1.00		
healthexp	0.23	1.00	
avgsch	0.79	0.31	1.00

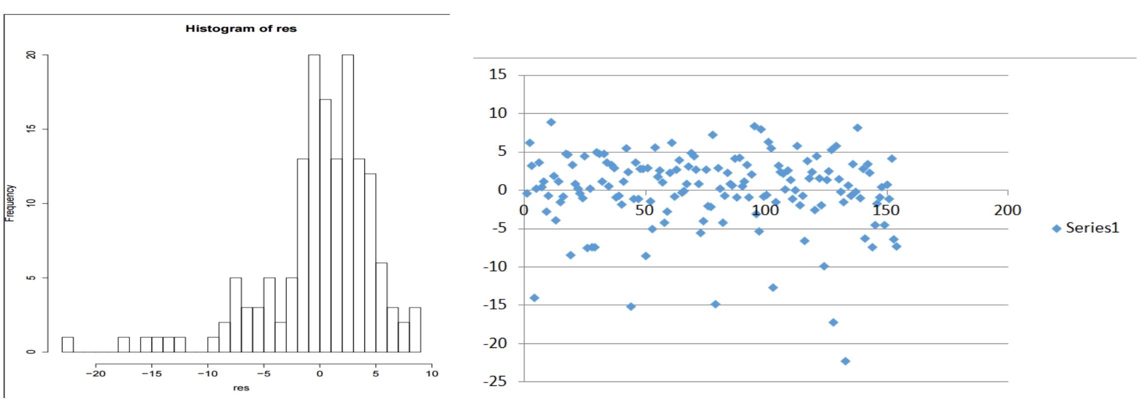
(Table 2: Correlation between explanatory variables)

4.2 Statistical Inference

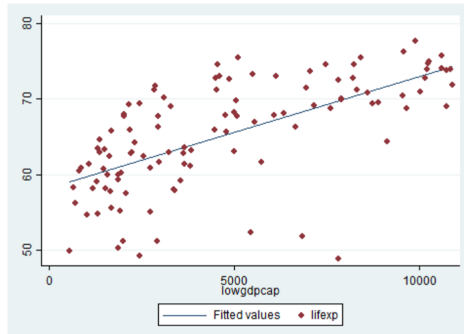
For statistical inference, T-test and F-test are checked for Regression Model 4. T-test is for understanding the significance of a coefficient; on the other hand, F-test checks all the slope coefficients and their statistical significance. For the T-test, the null hypothesis is $\beta = 0$; alternative hypothesis is $\beta \neq 0$, for all coefficients. As seen in the STATA table, the t-statistic for the coefficient for log(gdp per capita) is 7.18, which has a p-value of 0%. Because of this infinitesimal value, the null hypothesis is rejected and $\beta \neq 0$. The coefficient of public health expenditure has a t-statistic of 2.62 and a p-value of 1%. For the coefficient for average years of schooling, the t-statistic is 3.67, which also has a p-value of 0%. The null hypothesis is rejected for all three coefficients, and all explanatory variables are statistically significant at a 5% significance level. They should be included in the model.

For a multiple regression analysis, an F-test is also performed. For the t-test, each slope coefficient, β1, β2 and β3 are individually checked for statistical significance. With the F-test, the joint significance of the slope coefficients is tested, and the variance inflation factor skewing the individual t-statistics becomes irrelevant. The F-test gives a better idea of if every explanatory variable should be in the model. The null hypothesis states that β1=β2= β3=0; the alternative hypothesis says that the null isn't true and the variables are jointly significant. When performing the F-test, the F-statistic is 106.88, with a p-value of 0%. Because of this microscopic p-value, the null hypothesis is rejected. The three explanatory variables are jointly significant, so they all must be in the regression equation, confirmed once again.

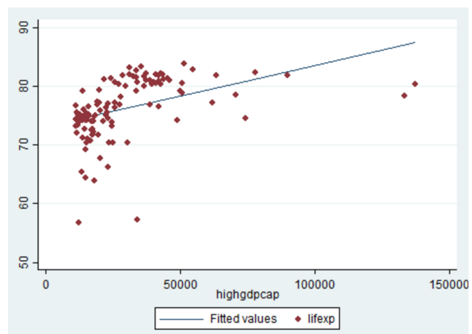
To use this model, the five Gauss-Markov Assumptions must also be checked. The first assumption states that the model is linear in parameters, in the betas. All of the coefficients, β0 to β3 are linear. Assumption 2 states that a random sample is used. All the data comes from public sources, which confirm they are random. Assumption 3 states that there is no perfect collinearity, which is already proven in Table 2 in Regression Model 4. Assumption 4 is the Zero Conditional Mean assumption: the expected values of the residuals should be 0. This assumption is met, as seen in the histogram below. The histogram is approximately Normal, centered at 0. Finally, the fifth Gauss-Markov Assumption is that the variance of the error terms is constant. As seen in the scatter plot of the residuals below, they all seem to fall in a line, with approximately constant variance. There are a few outliers, but overall, the variance looks very constant.



(Figure 5: Histogram and scatter plot of residuals)



To deeper analyze the data, a regression on the below-median GDP per capita and life expectancy was performed. Note that this is GDP per capita, not log(GDP per capita). The purpose of this regression is to analyze the curve in Regression Model 1. This analysis only included 114 observations but it did result in a R-squared value of 42.13%, which is higher than the R-squared for Regression Model 1. In addition, the p-value is 0% again, further emphasizing that β1 ≢ 0; there is a correlation between below-average GDP per capita and life expectancy. The regression for the below average GDP per capita and life expectancy is shown in the Appendix.



It is most important to note that this shows a slightly stronger correlation than the original, total data did.

Similarly, a deeper analysis was performed to look more closely at the relationship between above-median GDP per capita and life expectancy. This subset of data also only included 115 observations, and although the p-value is 0% it produced a low R-squared value of 19.11%. This is drastically smaller than the original R-squared value with all the observations of all GDP per capita, which had an R-squared of 37.78%. The summary for above-median GDP per capita and life expectancy is seen in the Appendix, as well.

5. Conclusion

To conclude, according to research, and as stated in the literature review, GDP per capita affects life expectancy up until a certain threshold. After some GDP, the correlation between the variables weakens. This analysis backs the claim. Our below-median GDP to life expectancy regression is stronger than the above-median GDP to life expectancy regression. In addition, when looking at the plot of overall GDP per capita and life expectancy, there is a clear curve shape. As the GDP per capita gets higher, the slope decreases. It affects the life expectancy less. This supports the claim that GDP per capita is more significant up until a certain maximum GDP point.

Dependent Variable: lifexp (life expectancy at birth, 2013)				
	Model 1	**Model 2**	**Model 3**	**Model 4**
gdpcap	0.00026 (11.74)***			
loggdpcap		5.74 (20.73)***	5.38 (17.21)***	3.84 (7.18)***
healthexp			0.33 (3.92)***	0.25 (2.62)***
avgsch				0.79 (3.67)***
intercept	66.29 (115.29)***	18.09 (7.06)***	17.63 (6.29)***	26.34 (6.85)***
# of observations	229	229	185	154
R-squared	37.78 %	65.43 %	67.35 %	68.13 %

***= significant at 1% **(Table 3: Summary of Models)**

Although the t-statistic has a probability of 0, implying that there is some relationship between above-median GDP per capita and life expectancy, the relationship is not that strong. It is not as strong as the relationship between below-median GDP per capita and life expectancy.

Therefore, in order to find a relationship between GDP per capita and life expectancy, some extra steps must be taken. The strongest model in predicting national life expectancy is when GDP per capita undergoes a log transformation and public health expenditure as a percentage of total government expenditure and average years of schooling are added as an additional explanatory variable. This can cancel the effects of the GDP per capita threshold, and provide a meaningful model for predicting life expectancy for a nation of any GDP per capita.

The data also provides a number of answers as well as future ideas. For example, the R-squared value is not very strong. Also, the estimators could be biased if under simplifying has occurred. To strengthen the regression as a whole, more explanatory variables can be added. In this way, R-squared value will increase, but in order to have a meaningful comparison between models with different number of explanatory variables, R-squared value should be used.

References

Becker, Gary S., Tomas J. Philipson, and Rodrigo R. Soares. "The Quantity And Quality Of Life And The Evolution Of World Inequality." American Economic Review 95.1 (2005): 277. Advanced Placement Source. Web. 01 Nov. 2016.

Deaton, A. "Health in an Age of Globalization." Brookings Trade Forum, vol. 2004 no. 1, 2004, pp. 83-130. Project MUSE, doi:10.1353/btf.2005.0004.

Deaton, Angus. "Income, Health, And Well-Being Around The World: Evidence From The Gallup World Poll." Journal Of Economic Perspectives 22.2 (2008): 53. Advanced Placement Source. Web. 01 Nov. 2016.

Hazan, Moshe. "Life Expectancy And Schooling: New Insights From Cross-Country Data." *Journal Of Population Economics* 25.4 (2012): 1237. *Advanced Placement Source*. Web. 16 Nov. 2016.

Appendix

Regression Model 1: Predicted Life Expectancy at birth = 66.28958 + 0.0002577(GDP per capita)

5 . regress lifexp gdpcap

Source	SS	df	MS
Model	5852.9176	1	5852.9176
Residual	9639.21503	227	42.4635023
Total	15492.1326	228	67.9479501

Number of obs	=	229
F(1, 227)	=	137.83
Prob > F	=	0.0000
R-squared	=	0.3778
Adj R-squared	=	0.3751
Root MSE	=	6.5164

Final Wednesday November 9 19:43:52 2016 Page 2

lifexp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdpcap	.0002577	.000022	11.74	0.000	.0002145	.000301
_cons	66.28958	.5749711	115.29	0.000	65.15662	67.42254

Regression Model 2: Predicted Life Expectancy = 18.09359 + 5.737335 x log(GDP per capita)

21 . regress lifexp loggdpcap

Source	SS	df	MS
Model	10135.9427	1	10135.9427
Residual	5356.1899	227	23.5955502
Total	15492.1326	228	67.9479501

Number of obs	=	229
F(1, 227)	=	429.57
Prob > F	=	0.0000
R-squared	=	0.6543
Adj R-squared	=	0.6527
Root MSE	=	4.8575

lifexp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
loggdpcap	5.737335	.2768173	20.73	0.000	5.191875	6.282795
_cons	18.09359	2.561391	7.06	0.000	13.04644	23.14073

Regression Model 3: Predicted Life Expectancy = 17.62939 + 5.384642 x log(GDP per capita) 0.3269995 x public health expenditure

9 . regress lifexp loggdpcap healthexp

Source	SS	df	MS
Model	9199.44432	2	4599.72216
Residual	4460.33512	182	24.5073358
Total	13659.7794	184	74.2379318

Number of obs	=	185
F(2, 182)	=	187.69
Prob > F	=	0.0000
R-squared	=	0.6735
Adj R-squared	=	0.6699
Root MSE	=	4.9505

lifexp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
loggdpcap	5.384642	.3128029	17.21	0.000	4.767455	6.001828
healthexp	.3269995	.0833462	3.92	0.000	.1625504	.4914486
_cons	17.62939	2.802131	6.29	0.000	12.10054	23.15823

Regression Model 4: Predicted Life Expectancy = 26.34697 + 3.843959 x log(GDP per capita) + 0.2454119 x public health expenditure + 0.7925244 x average years of schooling

```
8 . regress lifexp loggdpcap healthexp avgsch

      Source |       SS           df       MS      Number of obs   =       154
-------------+----------------------------------   F(3, 150)       =    106.88
       Model |  7950.32579         3  2650.1086    Prob > F        =    0.0000
    Residual |  3719.23926       150  24.7949284   R-squared       =    0.6813
-------------+----------------------------------   Adj R-squared   =    0.6749
       Total |  11669.5651       153  76.271667    Root MSE        =    4.9795

------------------------------------------------------------------------------
      lifexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   loggdpcap |   3.843959   .5350164     7.18   0.000     2.786817      4.9011
   healthexp |   .2454119   .0935534     2.62   0.010     .0605594    .4302645
      avgsch |   .7925244   .2161023     3.67   0.000     .3655267    1.219522
       _cons |   26.34697   3.847994     6.85   0.000      18.7437    33.95024
------------------------------------------------------------------------------
```

Regression Model of Life Expectancy and Below-Median GDP Per Capita

```
8 . regress lifexp lowgdpcap

      Source |       SS           df       MS      Number of obs   =       114
-------------+----------------------------------   F(1, 112)       =     81.52
       Model |  2387.30779         1  2387.30779   Prob > F        =    0.0000
    Residual |  3279.87674       112  29.2846137   R-squared       =    0.4213
-------------+----------------------------------   Adj R-squared   =    0.4161
       Total |  5667.18453       113  50.1520754   Root MSE        =    5.4115

------------------------------------------------------------------------------
      lifexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   lowgdpcap |    .001478   .0001637     9.03   0.000     .0011537    .0018024
       _cons |   58.20692   .9305715    62.55   0.000     56.36311    60.05073
------------------------------------------------------------------------------
```

Regression Model of Life Expectancy and Above-Median GDP Per Capita

```
8 . regress lifexp highgdpcap

      Source |       SS           df       MS      Number of obs   =       115
-------------+----------------------------------   F(1, 113)       =     26.70
       Model |  560.874088         1  560.874088   Prob > F        =    0.0000
    Residual |  2373.95988       113  21.0084945   R-squared       =    0.1911
-------------+----------------------------------   Adj R-squared   =    0.1840
       Total |  2934.83397       114  25.7441576   Root MSE        =    4.5835

------------------------------------------------------------------------------
      lifexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  highgdpcap |   .0001048   .0000203     5.17   0.000     .0000646     .000145
       _cons |   73.09672   .7409444    98.65   0.000     71.62878    74.56466
------------------------------------------------------------------------------
```